

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Ref. 3

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-020744

(43)Date of publication of application : 21.01.2000

(51)Int.Cl.

G06T 13/00
G06F 17/30

(21)Application number : 10-190185

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>

(22)Date of filing : 06.07.1998

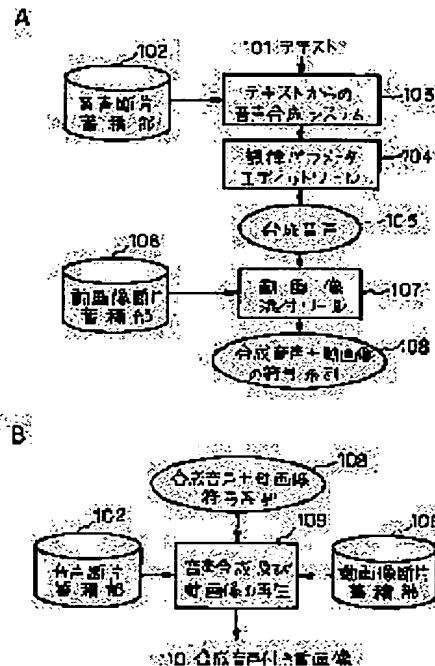
(72)Inventor : ABE MASANOBU

(54) METHOD FOR PRODUCING CONTENTS BY MOVING IMAGE AND SYNTHESIZED VOICE AND PROGRAM RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To attain movies (simultaneous reproduction of a voice and a moving picture) by small information volume.

SOLUTION: In the contents producing method, a voice is synthesized (103) from an input text 101 by a rule synthesis method by using voice fragments stored in a storage part 102, a contents producer constitutes a moving image by combining the moving image fragments of basic action (speaking, suprising and crying action, etc.), of a character which are stored in a storage part 106 in accordance with the contents of the text 101 and produces contents by respective indexes of these voice fragments and moving image fragments and a decoding side executes voice synthesis and the reproduction of a moving image by using voice fragments and moving image fragments corresponding to the indexes of the contents.



LEGAL STATUS

[Date of request for examination]

18.05.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-20744

(P 2 0 0 0 - 2 0 7 4 4 A)

(43) 公開日 平成12年1月21日 (2000.1.21)

(51) Int. Cl. ⁷	識別記号	F I	テーマコード [*] (参考)		
G06T 13/00		G06F 15/62	340	A	5B050
G06F 17/30		15/40	370	G	5B075
		15/403	330	Z	

審査請求 未請求 請求項の数 6 O L (全 4 頁)

(21) 出願番号 特願平10-190185

(22) 出願日 平成10年7月6日 (1998.7.6)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72) 発明者 阿部 匡伸

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74) 代理人 100066153

弁理士 草野 卓 (外1名)

Fターム(参考) 5B050 BA06 EA19 EA24 FA02 FA10

5B075 ND02 ND12 ND14 ND16 ND23

NK02 NK13 NK50

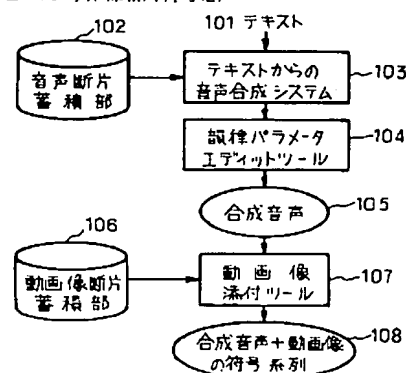
(54) 【発明の名称】 動画像と合成音声によるコンテンツ制作方法およびプログラム記録媒体

(57) 【要約】

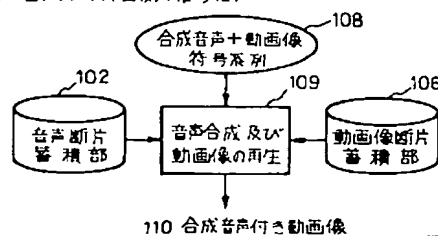
【課題】 少ない情報量でムービー（音声と動画を同時再生）を可能とする。

【解決手段】 入力テキスト101を、蓄積部102の音声断片を用いて規則合成法により音声合成し（103）、コンテンツ制作者はテキストの内容に応じて蓄積部106中の、キャラクタの基本動作（喋っている動作、驚きの動作、泣いている動作など）の動画像断片を組合せて、動画像を構成し、これらの音声断片、動画像断片の各インデックスにてコンテンツとし、復号側では、このコンテンツのインデックスで対応音声断片、動画像断片を用いて音声合成と動画像の再生を行う。

A コンテンツ作成側（符号化）



B コンテンツ再生側（復号化）



【特許請求の範囲】

【請求項 1】 音声の断片とそれを示すインデックス、および、動画像の断片とそれを示すインデックスとを蓄積しておき、

入力されたテキストに応じて前記蓄積音声断片を参照して音声断片のインデックス列を作ると共に韻律パラメータを付加し、かつ前記入力テキストに応じて前記蓄積動画像断片から選んで動画像断片のインデックス列を構成し、1つのコンテンツを作成する動画像と合成音声によるコンテンツ制作方法。

【請求項 2】 前記音声の断片として、規則合成方法で利用される音声断片を用い、規則合成方法によって前記符号列のテキストから音声合成を行い、その合成音声を用いることを特徴とする請求項 1 記載の動画像と合成音声によるコンテンツ制作方法。

【請求項 3】 前記音声の断片として、規則合成方法で利用される音声断片を用い、合成音声の韻律パラメータや、音素パラメータを制作者が操作して作成した合成音声を用いることを特徴とする請求項 2 記載の動画像と合成音声によるコンテンツ制作方法。

【請求項 4】 前記動画像の断片として、キャラクタの基本動作を蓄積しておき、この基本動作の組み合わせによって動画像の断片のインデックス列を構成することを特徴とする請求項 1 乃至 3 の何れかに記載の動画像と合成音声によるコンテンツ制作方法。

【請求項 5】 入力テキストに応じた音声と動画像を組合せたコンテンツを、コンピュータ処理により制作するプログラムを記録した記録媒体であって、前記入力テキストを解析して音声断片蓄積部の音声断片を用いて規則合成法によって音声合成してそのインデックス列として前記音声表現する処理と、前記入力テキストの内容に応じて、動画像断片蓄積部の動画像断片が選択合成されたものを、そのインデックス列として前記動画像を表現する処理と、前記音声断片のインデックス列及び前記動画像断片のインデックス列の組を 1つのコンテンツとして出力する処理とを上記コンピュータに行わせることを特徴とするプログラムを記録した記録媒体。

【請求項 6】 音声断片のインデックス列と動画像断片のインデックス列の組合せよりなるコンテンツをコンピュータにより再生するプログラムを記録した記録媒体であって、前記音声断片のインデックス列の各インデックスを、音声断片蓄積部を参照して音声断片列よりなる音声に変換する処理と、前記動画像断片のインデックス列の各インデックスを、動画像断片蓄積部を参照して動画像断片列よりなる動画像に変換する処理とを同時にコンピュータに実行させることを特徴とするプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、ムービーのように動画像と音声とを同時に表示できるコンテンツを制作する方法及びプログラム記録媒体に関する。

【0002】

【従来の技術】 物事の意味や、物の使い方を説明する場合、文字、図、写真等だけによる説明、または、音声だけによる説明では、理解者にとっては、理解が困難であったり、理解に時間を要することがまある。また、解説者にとっては、解説するための資料の作成などに多大な労力が必要となる。これに対し、音声と動画像、音声とアニメーションを組み合わせれば、理解し易い説明を容易に実現できることが多い。また、音声だけによるアナウンスよりも、キャラクタが身振り手振りをしながら音声でアナウンスする場合の方が、面白く、親しみが湧くことがある。この発明によれば、以上のようなコンテンツを作成することができる。

【0003】 従来のムービーは、人間や背景の録画、コンピュータによる動画像の制作、アニメーションによる動画の作成と、人間の発声した音声や効果音との組み合わせで構成されている。これらのムービーは、情報量が多く、その蓄積や伝送のための費用が高い。

【0004】

【発明が解決しようとする課題】 上述のように、ムービーなどのような音声と動画像を同時に利用できるコンテンツは便利なものであるが、コンテンツの容量が多く、それらのコンテンツを快適に利用するためには制限がある。特に、最近広く普及したインターネットは、伝送容量が少ないため伝送時間が長くなり、ムービーを使うのに適した環境であるとは言い難い。この発明は、ムービーをネットワーク経由で利用する上で、その伝送量が膨大であるという問題を解決する。

【0005】

【課題を解決するための手段】 日本語を構成する音素は有限であり、任意の日本語は有限個の音素（音声の断片）の組み合わせで表現できると考えられる。また、ムービーであっても、ある限られたタスクにおいては、有限個の動画像の断片を組み合わせで表現することが可能である。この発明では、有限個の音声の断片や有限個の動画像の断片を蓄積しておき、これらを組み合わせることによってムービーを音声断片や動画像断片のインデックス列として再構成する。これによって、伝送側から受信側に送られるデータは、音素や動画像の断片ではなく、そのインデックス情報で済み、格段に伝送容量を減らすことができる。

【0006】

【発明の実施の形態】 図 1 はこの発明の一実施例の構成を示す。図 1 A はコンテンツを作成する流れを、図 1 B は作成したコンテンツを再生する流れを示す。これを参照して具体的動作について説明する。コンテンツを作成

10

20

30

40

50

する手順において、入力 101 は、ワープロ等で作成したかな漢字混じり文、つまりテキストである。音声断片データが音声蓄積部 102 に蓄積されている。この実施例では、任意の日本語を合成する場合を想定しているので、音声蓄積部 102 に蓄積されている音声断片データは日本語の音素や音素を連ねた音声断片とそのインデックスを集めたものである。任意の日本語を合成する必要が無い場合には、限られた音声とそのインデックスを蓄えておくことも考えられる。

【0007】音声合成器 103 は、入力されたテキスト 101 を、音声蓄積部 102 を参照して音声合成する。合成音声における基本周期、長さ、強さなどの韻律パラメータを操作するためのツール 104 は音声合成器 103 で合成された音声のイントネーションが不十分であると考えられるときにコンテンツの制作者が使うものである。一般に、テキストから音声合成する場合、感情や文脈等の情報を適切に表現することが難しいため、韻律パラメータエディットツール 104 によるエディットは有効である。以上の結果として合成音声 105 が生成される。なお、この種のエディットツールの手法については、例えば情報処理学会音声言語情報処理研究会 1997 年 7 月 19 日発表の予稿「様々な音声表現を実現できる音声作成ツール」に記載されている。

【0008】動画像蓄積部 106 に蓄積する動画像断片データとして、キャラクタの基本動作とそのインデックスを蓄積しておくことが考えられる。キャラクタの基本動作とは、前向きで喋っている動作、横向きで喋っている動作、驚きの動作、泣いている動作、誉めている動作、走っている動作などであり、これらの基本動作を組み合わせることによってキャラクタに演技をさせる。ツール 107 で、合成音声 105 に対して、動画像断片データを添付する。つまり合成音声 105 の部分部分に、その発話内容に適した基本動作を制作者が動画像蓄積部 106 から選択して割り当てる。さらに、合成音声 105 の開始時間、動画像断片の開始時間の相対関係を設定したり、合成音声と動画像の時間長が異なる場合の動作を設定する。例えば、合成音声が動画像より短い場合には、音声の終了と同時に動画像を中断するか、音声終了しても動画像を最後まで再生するか、を設定する。逆に、合成音声が動画像より長い場合には、動画像を繰り返し再生するか、動画像を終了時点で停止させておく

か、を設定する。以上のように、合成音声 105 と動画像との対応付けを設定した後に、これら合成音声と動画像について音声の断片インデックス（符号）、音声合成のための韻律情報、動画像断片インデックス（符号）および、合成音声と動画像の時間関係の情報を保存し、符号系列、つまりコンテンツ 108 とする。このコンテンツ 108 に保存されるデータは、これらを用いて再生される動画像、および、音声に比べると 4 桁以上少なくなる。

【0009】次に、図 1B を参照して、コンテンツ 108 を再生する手順を説明する。図 1A に示したコンテンツ作成手順にしたがって作成されたデータ（コンテンツ）108 は、そのコンテンツ作成で用いた音声断片データを音声蓄積部 102 から、また動画像断片データを動画像蓄積部 106 からそれぞれ取出し、これらデータを基に、音声の合成と動画の再生を合成・再生部 109 で行って、ムービーを構成する。

【0010】なおコンテンツ制作時に、合成音声の韻律パラメータとしては、規則合成で得られたものをそのまま用い、エディットツール 104 による修正は行わなくてもよい。

【0011】

【発明の効果】この発明の特徴は、（１）ムービーを構成する動画像と音声を、有限の音声断片データと動画像断片データとによって表現すること、（２）コンテンツ作成で用いた音声断片データと動画像断片データとを、コンテンツ再生側に蓄積しておくこと、にある。その結果、コンテンツ作成側からコンテンツ再生側に送られるデータはその音声断片データ、動画像断片データの各インデックスとすることができ、再生されるムービーの情報量に比べて、格段に少ない情報量ですむ。

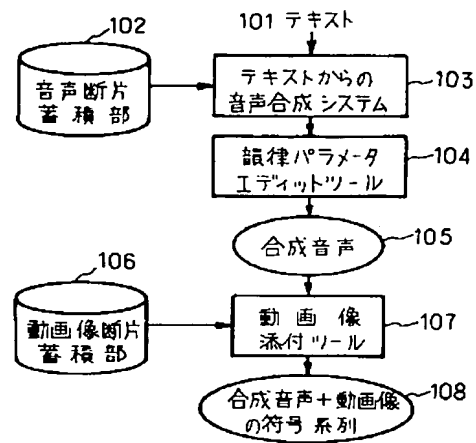
【0012】この発明によって作成されたコンテンツは、情報量が少ないため、インターネットなど、伝送速度が遅い場合でも待ち時間を少なく抑さえ、ムービーによる表現を実現できる。また、伝送速度が速い場合でも、安価にムービーを利用できることとなる。

【図面の簡単な説明】

【図 1】A はこの発明によるコンテンツ制作方法の処理手順の例を示す流れ図、B はそのコンテンツから合成音声付き動画像を作成する方法の手順を示す流れ図である。

【図 1】

A コンテンツ作成側(符号化)



B コンテンツ再生側(復号化)

